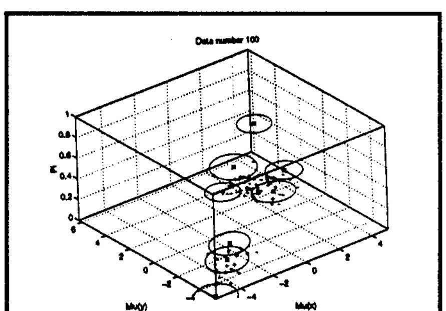
A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators

Jeffrey L. Solka
Wendy L. Poston
Edward J. Wegman

Technical Report No. 114
December, 1994

Center for Computational Statistics



George Mason University
Fairfax, VA 22030

CENTER FOR COMPUTATIONAL STATISTICS
RECENT TECHNICAL REPORTS

TR 97. Wendy L. Poston and Jeffrey L. Solka, A Parallel Method to Maximize the Fisher Information Matrix, June, 1994.

TR 98. Edward J. Wegman and Charles A. Jones, Simulating a Multi-target Acoustic Array on the Intel Paragon, June, 1994.

TR 99. Barnabas Takacs, Edward J. Wegman and Harry Wechsler, Parallel Simulation of an Active Vision Model, June, 1994.

TR 100. Edward J. Wegman and Qiang Luo, Visualizing Densities, October, 1994.

TR 101. Daniel B. Carr, Converting Tables to Plots, October, 1994.

TR 102. Julia Corbin Fauntleroy and Edward J. Wegman, Parallelizing Locally-Weighted Regression, October, 1994.

TR 103. Daniel B. Carr, Color Perception, the Importance of Gray and Residuals on a Choropleth Map, October, 1994.

TR 104. David J. Marchette, Carey E. Priebe, George W. Rogers and Jeffrey L. Solka, Filtered Kernel Density Estimation, October, 1994.

TR 105. Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe, Wendy L. Poston and George W. Rogers, A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap, October, 1994.

TR 106. Wendy L. Poston, Edward J. Wegman, Carey E. Priebe and Jeffrey L. Solka, A Contribution to the Theory of Robust Estimation of Multivariate Location and Shape: EID, October, 1994.

TR 107. Clifton D. Sutton, Tree Structured Density Estimation, October, 1994.

TR 108. Charles A. Jones, Simulating a Multi-target Acoustic Array on the Intel Paragon (M.S. Thesis), October, 1994.

TR 109. Leonard B. Hearne and Edward J. Wegman, Fast Multidimensional Density Estimation based on Random-width Bins, October, 1994.

TR 110. Edward J. Wegman, Huge Data Sets and the Frontiers of Computational Feasibility, November, 1994.

TR 111. Winston C. Chow, Fractional Process Modeling, November, 1994.

TR 112. Mark C. Sullivan, Computationally Efficient Statistical Signal Processing Using Nonlinear Operators (Ph.D. Dissertation), December, 1994.

TR 113. Irwin Greenberg, Some Simple Approximation Methods in Level Crossing Problems, December, 1994.

TR 114. Jeffrey L. Solka, Wendy L. Poston and Edward J. Wegman, A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators, December, 1994.

A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators

Jeffrey L. Solka[1], Wendy L. Poston[1], and Edward J. Wegman[2]

[1]Dahlgren Division of the Naval Surface Warfare Center
Systems Research and Technology Department
Advanced Computation Technology Group, Code B10
Dahlgren, VA 22448-5000

[2]Department of Statistics
George Mason University
Fairfax VA 22030-4444

Corresponding Author - Jeff Solka
703-663-1982 (W)
703-663-1952 (FAX)
703-663-7226 (SECY)
jsolka@relay.nswc.navy.mil (EMAIL)

This paper focuses on recent work which analyzes the expectation maximization (EM) evolution of mixtures based estimators. The goal of this research is the development of effective visualization techniques to portray the mixture model parameters as they change in time. This is an inherently high dimensional process. Techniques are presented which portray the time evolution of univariate, bivariate, and trivariate finite and adaptive mixtures estimators. Adaptive mixtures is a recently developed variable bandwidth kernel estimator where each of the kernels is not constrained to reside at a sample location. The future role of these techniques in developing new versions of the adaptive mixtures procedure are also discussed,

(Submitted to The Journal of Computational and Statistical Graphics 11/94)

Keywords - density estimation, expectation maximization, graphics

A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators

This paper focuses on recent work which analyzes the expectation maximization (EM) evolution of mixtures based estimators. The goal of this research is the development of effective visualization techniques to portray the mixture model parameters as they change in time. This is an inherently high dimensional process. Techniques are presented which portray the time evolution of univariate, bivariate, and trivariate finite and adaptive mixtures estimators. Adaptive mixtures is a recently developed variable bandwidth kernel estimator where each of the kernels is not constrained to reside at a sample location. The future role of these techniques in developing new versions of the adaptive mixtures procedure are also discussed,

1: Introduction

Given $X = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$ where each $\vec{x}_i$ is d-dimensional and i.i.d. according to an unknown density $f_0(\vec{x})$ one is often interested in estimating $f_0(\vec{x})$. This problem occurs in such areas as exploratory data analysis, classification, and regression. There are a variety of approaches to the multivariate density estimation problem (Scott, 1992).

An often used parametric approach is that of finite mixtures density estimation (FMDE) (Everitt and Hand, 1981) in combination with the expectation maximization (EM) method of Dempster, Laird, and Rubin (1977). One difficulty with this tactic is that one needs some idea as to the appropriate number of terms in the mixture model as well as the approximate parameter values. Given this information the EM algorithm is guaranteed to converge to at least a local maxima in the likelihood surface.

Some of the previous nonparametric approaches include histograms (Sturges, 1926), frequency polygons (Scott, 1985a), adaptive histograms (Wegman, 1970), average shifted histograms (Scott, 1985b), and kernel estimators (Silverman, 1986). These approaches are beneficial in that they possess nice asymptotic consistency properties, robustness with regard to nonnormality, and fewer parameters to estimate which implies better estimates in the finite sample regime. They are at a disadvantage as compared to the mixture model

approach when it is suspected that the unknown true density is a mixture of a number of terms and one would like to estimate the posteriori probability of underlying term membership for an unlabeled observation.

A recently developed density estimation technique that circumvents some of the problems of the above techniques is the adaptive mixtures density estimation (AMDE) procedure of (Priebe, 1994). This procedure is a blend of the finite mixtures and kernel estimator approaches. It is essentially a mixtures-type approach that allows for the creation of new terms in a data driven manner. We have successfully applied this technique in combination with fractal-based features to the detection of man-made objects in land (Solka, Priebe, and Rogers, 1992) and aerial (Priebe, Solka, and Rogers, 1993) images, the general problem of texture classification (Solka, Priebe, and Rogers, 1993), and the measurement of breast parenchymal tissue density (Priebe and Solka et al., 1994). The adaptive mixtures estimator is asymptotically consistent like the kernel estimator, but it has the added benefit of creating additional terms at a rate which is considerably less then the rate n creation associated with the kernel estimator.

An inherent difficulty with both the parametric FMDE and the nonparametric AMDE is understanding the time evolution of the system under the EM equations. Even in the simple FMDE case of a two component mixture the evolution of the parameters is a five-dimensional process. The situation is worse in the case of AMDE since the dimensionality of the problem increases each time a new term is added to the model. We will discuss within a new visualization technique that makes the problem of understanding this time evolution more tractable.

There are several reasons why the ability to monitor this time evolution is important. In the case of FMDE the nature of the likelihood surface that drives this evolution is

very poorly understood and it is hoped that new insights into the nature of the likelihood surface can be obtained through close monitoring of the evolution of the parameters. Second it is well known that the FMDE under the EM method is only guaranteed to converge to a local maxima in the likelihood equations (Redner and Walker, 1984). One usually circumvents this difficulty by starting the mixture model at a variety of initial conditions in parameter space. Our visualization technique provides a convenient way to monitor the process so that one can restart the procedure earlier.

In the case of AMDE even less is known about the behavior of the system. We have used our visualization techniques to help expand our understanding of not only the dynamics of the system, but also the character of the solutions that the procedure produces. This has led us to the more efficient formulation of alternative local bandwidth estimators. Last but not least we point out the known utility that visualization techniques provide with regard to software verification. It is much easier to validate the workings of a software system using visualization techniques in combination with analytical procedures.

In Section 2 of the paper we present a quick review of FMDE and AMDE. This is followed by discussions of some earlier attempts at visualization of AMDE models. We also present our new approach for the visualization of univariate, bivariate, and trivariate FMDE and AMDE. In Section 3 we present univariate, bivariate, and trivariate results obtained using this new visualization procedure. These results illustrate the utility of the procedure. Specific cases are presented that highlight some of the insights that can be obtained using the procedure. The Section goes on to explain how to obtain on-line access to the movies that detail the examples presented within. In the the final Section we sum-

marize the results presented and look ahead to future research efforts.

2: Approach

Finite Mixtures Density Estimation

Given an unknown distribution $f_0(\vec{x})$ we seek to model the distribution using $\hat{f}(\vec{x};\hat{\Psi})$ defined by

$$\hat{f}(\vec{x};\hat{\Psi}) = \sum_{i=1}^{g} \hat{\pi}_i K(\vec{x};\hat{\theta}_i) , \tag{1}$$

where K is some fixed density parameterized by $\hat{\theta}_i$, and $\hat{\Psi} = \left(\hat{\pi}_1, \hat{\theta}_1, \hat{\pi}_2, \hat{\theta}_2, ..., \hat{\pi}_g, \hat{\theta}_g\right)$. The $\hat{\pi}_i$""s are referred to as the mixing proportions. We can assume for much of what follows that K is taken to be the normal distribution, in which case $\hat{\theta}_i$ becomes $\{\hat{\mu}_i, \hat{\Sigma}_i\}$. In the simplest case the mixture is assumed to have a single term and the parameters that must be estimated are the mean and covariance of the distribution.

In the case of FMDE we begin with an initial guess as to g the number of components and $\hat{\psi}$ their parametric values. Given this initial "guess" $\hat{\psi}$ is updated based on the iterative EM equations as follows:

$$\hat{\tau}_{ij} = \frac{\hat{\pi}_i \hat{f}_i(\vec{x}_j;\hat{\theta})}{\sum_{t=1}^{g} \hat{\pi}_t \hat{f}_t(\vec{x}_j;\hat{\theta})} , \tag{2}$$

$$\hat{\pi}_{ij} = \sum_{j=1}^{n} \frac{\hat{\tau}_{ij}}{n} , \tag{3}$$

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} \hat{\tau}_{ij} \vec{x}_j}{n \hat{\pi}_i}, \text{ and} \tag{4}$$

$$\hat{\Sigma}_i = \sum_{j=1}^{n} \frac{\hat{\tau}_{ij}\left(\vec{x}_j - \hat{\mu}_i\right)\left(\vec{x}_j - \hat{\mu}_i\right)^{\mathrm{T}}}{n \hat{\pi}_i}. \tag{5}$$

This is where $\hat{\tau}_{ij}$ is the estimated posteriori probability that x_j belongs to term i, $\hat{\pi}_{ij}$ is the estimated mixing coefficient, $\hat{\mu}_i$ is the d-dimensional estimated mean vector, and $\hat{\Sigma}_i$ is the dxd estimated covariance matrix for the ith term.

Adaptive Mixtures Density Estimation

There is an alternate formulation of the EM update equations that recursively updates the estimate of the parameters $\hat{\Psi}$ based on a single new observation. This version provides the capability to update the parameter estimates without storage of the data set, but at the cost of much slower convergence. The AMDE was first formulated in terms of this recursive approach. The exact form of the update equations is as follows:

$$\hat{\tau}_{n+1}^{(i)} = \frac{\pi_n^{(i)} \hat{f}^{(i)} (\vec{x}_{n+1}; \hat{\theta}_n)}{\sum_{t=1}^{g} \pi_n^{(t)} \hat{f}^{(t)} (\vec{x}_{n+1}; \hat{\theta}_n)} \tag{6}$$

$$\hat{\pi}_{n+1}^{(i)} = \hat{\pi}_n^{(i)} + \frac{1}{n}\left(\hat{\tau}_{n+1}^{(i)} - \hat{\pi}_n^{(i)}\right) \tag{7}$$

$$\hat{\mu}_{n+1}^{(i)} = \hat{\mu}_n^{(i)} + \frac{\hat{\tau}_{n+1}^{(i)}}{n \hat{\pi}_n^{(i)}}\left(\vec{x}_{n+1} - \hat{\mu}_n^{(i)}\right), \text{ and} \tag{8}$$

$$\hat{\Sigma}_{n+1}^{(i)} = \hat{\Sigma}_n^{(i)} + \frac{\hat{\tau}_{n+1}^{(i)}}{n\hat{\pi}_n^{(i)}}\left[\left(\vec{x}_{n+1} - \hat{\mu}_n^{(i)}\right)\left(\vec{x}_{n+1} - \hat{\mu}_n^{(i)}\right)^T - \hat{\Sigma}_n^{(i)}\right]. \qquad (9)$$

Here $\hat{\tau}_{n+1}^{(i)}$ is the estimated posteriori probability of $\vec{x}_n$ belonging to the ith term of the mixture, $\hat{\pi}_{n+1}^{(i)}$ is the estimated mixing coefficient, $\hat{\mu}_{n+1}^{(i)}$ is the d-dimensional estimated mean, and $\hat{\Sigma}_{n+1}^{(i)}$ is the dxd estimated covariance matrix of the ith term.

The AMDE stochastic approximation approach is to recursively update $\hat{\Psi}$, the estimate of the true parameters Ψ_0, while simultaneously providing the capability to expand the parameter space $\hat{\Psi}$ if dictated by the complexity of the data. We note that in the AMDE case, the parameter space $\hat{\Psi}$ is given by $\hat{\Psi} = \left(\hat{\pi}_1, \hat{\theta}_1, \hat{\pi}_2, \hat{\theta}_2, \ldots\right)$. The procedure

$$\hat{\Psi}_{t+1} = \hat{\Psi}_t + A \cdot U_t(\vec{x}_{t+1};\hat{\Psi}_t) + B \cdot C_t(\vec{x}_{t+1};\hat{\Psi}_t, t) , \qquad (10)$$

is used to recursively update the density. Here $A = \left[1 - P_t(\vec{x}_{t+1};\hat{\Psi}_t)\right]$ and $B = P_t(\vec{x}_{t+1};\hat{\Psi}_t)$. P_t represents a possibly stochastic create decision and takes on values 0 or 1. U_t updates the current parameters using equations (6-9) while C_t adds a new term to the model. As is implicit in the equation, the decision to add a new term is a function of the current data point, our current estimation of the parameters, and time. The time dependence is important in those cases where we wish to anneal the probability of creation as a function of training time.

The exact nature of the creation process is as follows. The Mahalanobis distance from the new observation $\vec{x}_t$ to each of the terms in the model is computed using $MHD(i) = \left(\vec{x}_t - \hat{\mu}^{(i)}\right)^T \hat{\Sigma}^{-1(i)}\left(\vec{x}_t - \hat{\mu}^{(i)}\right)$. If MHD(i)>τ_c (a create threshold) for every term then a new term is created at $\vec{\mu}^{(new)} = \vec{x}_t$, with a covariance given by $\Sigma^{(new)} = \Im\left(\Sigma^{(i)}\right)$ and a mixing coefficient of $\hat{\pi}^{(new)} = \frac{1}{n}$ assuming $\vec{x}_t$ is the nth data point. $\Im(.)$ is a weighted average based on posteriori probability. We also note that the

mixing coefficients of the remaining terms are all equally decremented to accommodate the new term.

dF Space Representation

The challenge is to design an effective visualization technique to monitor the evolutions of systems under either the finite or adaptive mixtures process. As indicated previously, the technique needs to deal with the inherent high dimensionality of the problem. In addition it needs to provide a realistic portrayal of the system.

We have previously presented one attempt at static visualization of adaptive mixtures models (Priebe and Lorey et al, 1994). This approach came as a natural by-product of casting a mixture model within a Bayesian framework. In this case, we can write our estimate as

$$\hat{f}(x) = \int_{\Omega} N(\mu, \sigma)\, dF \,, \qquad (11)$$

where dF is the measure for the parameter space. In the case of discrete mixtures dF becomes a probability mass distribution and the integral is converted into a sum. We may represent the distribution associated with dF as a group of points in (μ,σ^2,π) space. Our previous work rendered the support of dF in R^2 by plotting a circle whose radius is determined by the term's mixing coefficient and whose center is given by the term's mean and variance. For example we represent the two component mixture $f_0(x)=.5*N(-$

2,.1)+.5*N(2,1) as follows, please see Figure 1.

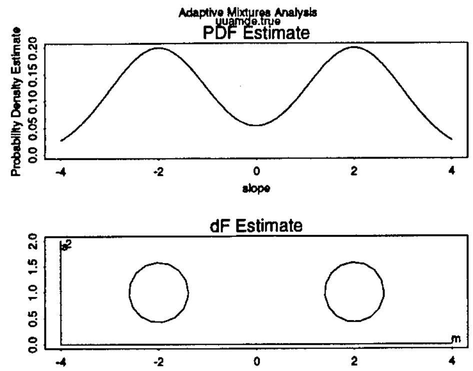


Figure 1:dF space representation of $f_0(x)$=.5*N(-2,.1)+.5*N(2,1).

While this approach has the advantage of truly representing the support of the underlying parametric distribution function there is no convenient way to extend it to bivariate and trivariate mixtures. With this end in mind we propose the following approach.

Univariate Representation

In the case of univariate mixtures we represent each term in the mixture as a magenta ellipse whose major radius is related to the standard deviation of the term and whose center is given by the mean and mixing coefficient. The graphics device can make the ellipsoids appear somewhat circular. In addition since our goal is monitoring the development of the model as influenced by the data, we also propose to include functional plots of the underlying density function from which the data was drawn when available as a green line and the current mixture model as a magenta line along with a scatter plot of the data set along the μ_x axis. Returning to our consideration of the radius of the term we have chosen to set the radius exactly equal to the standard deviation of the term. In Figure

2 we present this representation of $f_0(x)=.75*N(-2,.25)+.25*N(2,2)$ as the true probability density function with the current state of the mixture model at $\hat{f}(x) = .5*N(-2,.1)+.5*N(2,1)$.

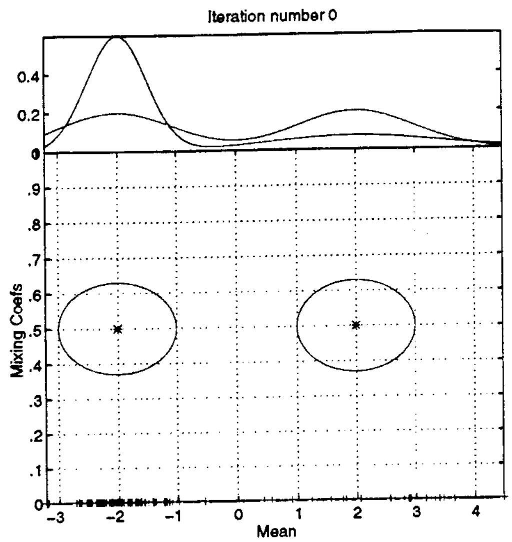


Figure 2: Sample screen snapshot for the univariate em algorithm case.

In the case of adaptive mixtures we employ the added convention of indicating the latest term created by using a red '+' instead of a '*' at its center.

Bivariate Representation

We next discuss the bivariate case which follows naturally once we step away from the dF space representation. We represent each term in the mixture as an ellipse whose eccentricity is determined by the solution of $\left(\vec{x}-\hat{\mu}^{(i)}\right)^T \Sigma^{(i)-1}\left(\vec{x}-\hat{\mu}^{(i)}\right) = 1$. Hence we represent the term as a magenta ellipse centered at $(\mu^{(i)}_x, \mu^{(i)}_y, \pi^{(i)})$ which resides in (μ_x, μ_y, π) space and is parallel to the (μ_x, μ_y) plane. As before we form a scattter plot of the data in the (μ_x, μ_y) plane. Figure 3 provides an example plot based on a bimodal two

component mixture whose structure is given by

$$\hat{f}(x,y) = 0.5N\left((-3,-3),\begin{bmatrix}1.0 & 0.8\\ 0.8 & 1.0\end{bmatrix}\right)+0.5N\left((3,3),\begin{bmatrix}1.0 & -0.8\\ -0.8 & 1.0\end{bmatrix}\right) \tag{12}$$

We have included 100 points drawn from $f_0 = \hat{f}$ as part of the illustration.

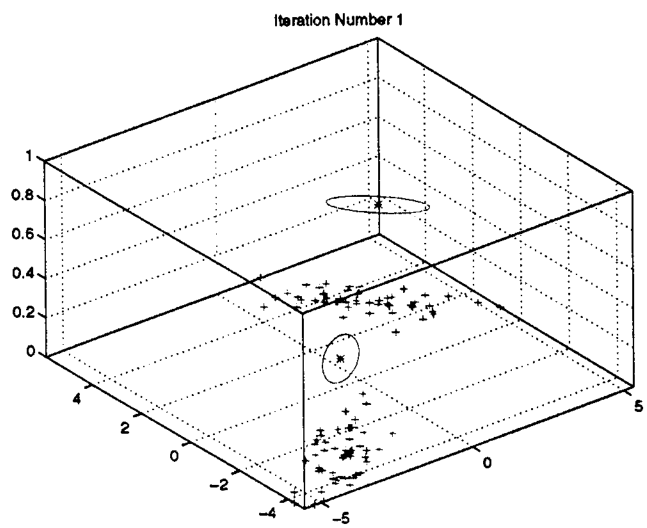


Figure 3: Sample screen snapshot for the bivariate finite mixtures case.

Once again we indicate the presence of a new term by using a '+' rather then a '*' at the center of the ellipse.

Now that we have made the transition into three space a quick word about viewpoints is in order. We follow the MATLAB convention and specify our viewpoint as a two

vector (ϕ,θ) where ϕ is the rotation angle about the z axis measured in degrees where positive angles (where 0 coincides with the x-axis) represent counter clockwise rotation and θ is the elevation angle of the viewing eye measure with respect to the xy plane in degrees. The viewpoint in Figure 3 is the default viewpoint of (-37.5,45).

Trivariate Representation

In this case each term is plotted as an ellipsoid in (μ_x,μ_y,μ_z) space. The ellipsoid is determined by $\left(\vec{x}-\hat{\mu}^{(i)}\right)^T \Sigma^{(i)^{-1}}\left(\vec{x}-\hat{\mu}^{(i)}\right) = 1$. So each "egg" is plotted at $(\mu^{(i)}_x,\mu^{(i)}_y,\mu^{(i)}_z)$. Since the trivariate nature of the term fully occupies the underlying dimensionality of the embedding space, we are faced with the question of how to represent the mixing coefficient for this term. We have chosen to use the color of the egg to indicate each term's mixing coefficient. We present the color ramp, mappings from colors to π's, above the plot for ease of reference by the user. We have chosen not to scatter plot the underlying data in this case so as not to clutter the graph. Figure 4 presents a plot of the mixture $\hat{f}(x, y, z) = 0.5N((-4,-4,-4), I_d) + 0.5N((4,4,4), I_d)$ where I_d is the 3x3 identity matrix.

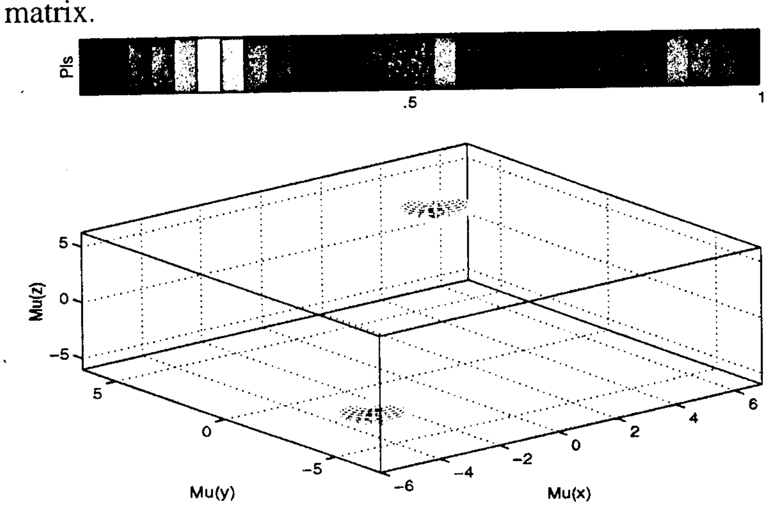


Figure 4: Trivariate two term test case.

3: Results

MATLAB Implementation

The visualization code was initially developed on a 486/33MHz computer using MATLAB 4.2. The code was then transferred to a SILICON GRAPHICS INDY 2 platform for further development. MATLAB was chosen because of it's computational capabilites as well as it's many graphics tools; e.g., the ability to make movies of the density estimation process. There is nothing in the code or the process that makes it machine dependent, which allows for a wider usage. Additionally, the authors took care to use only those functions that come with the MATLAB package itself; i.e, no toolbox functions were used in the implementation. The functions are written in a modular manner for greater adaptability and ease of use. Several switches are implemented that allow the user to tailor a given run. For example, a user may want to run FMDE without graphical output or print screen snapshots at certain iterations.

Univariate Results

We present results that illustrate the application of the procedure to univariate, bivariate, and trivariate finite and adaptive mixtures models. Each test case has been chosen to best illustrate the effectiveness of the procedure. As can be expected it is difficult to display what is a dynamic process in a set of stills. It is hard to fully appreciate the process without the use of movies. We will have more to say about the subject of movies at the end of this section.

The first test case consists of 1000 points drawn from the mixture $f_0(x) = .25N($-

6,1) + .25N(-2,1)+.25N(2,1)+.25N(6,1). We illustrate our technique by considering the evolution of a 4 component finite mixture model under this data set. The initial settings of the model are as follows:

π_1=.05,μ_1=-10,σ_1^2=1.3;

π_2=.05,μ_2=-5,σ_2^2=.03;

π_3=.45,μ_3=0,σ_3^2=.03;

π_4=.53,μ_4=10,σ_4^2=1.3.

This initial model is displayed in Figure 5.

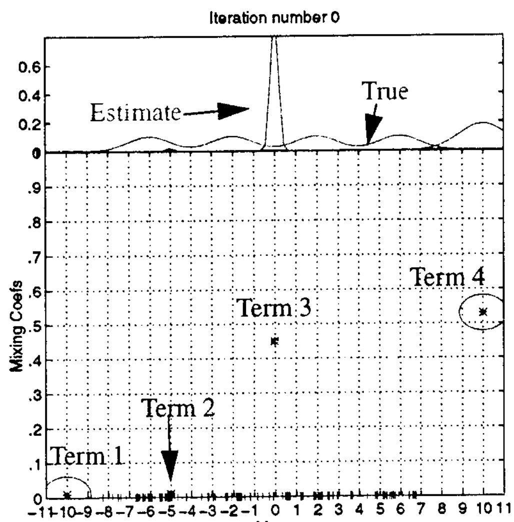


Figure 5: Initial configuration of the 4 mode 4 term finite mixture test case.

The top frame contains a standard functional representation of the probability density functions for the mixture model rendered in magenta and for the true model rendered in green . In the bottom frame each term in the model takes the form of an ellipse and the first 100 points of the data is plotted in green along the x-axis. The initial configuration of the x axis is data driven and this in part leads to only a partial display of the initial terms. As is expected by the nature of the EM algorithm the terms are ultimately drawn into a more

close interaction with the data and hence this display problem is solved.

Figure 6 displays the model after the first iteration through the data. We notice that

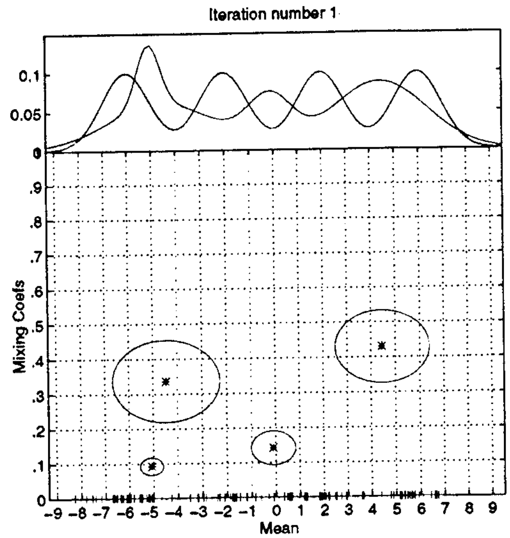


Figure 6: The 4 mode 4 term finite mixtures test case after the first iteration through the data.

there has been a large adjustment on the parameters at the end of the first step. This is not surprising given how far "off track" the parameters initially started. As will be seen, subsequent frames will indicate that this initial adjustment is much larger then the later ones and is suggestive of the steep nature of the likelihood surface at the perimeter. These types of insights are one of the benefits of the visualization process. Figures 7 a, b, and c portrays the solution at 10, 25, and 50 iterations respectively. The final parameters in the model are given by:

π_1=.2476,μ_1=-2.1143,σ_1^2=1.1239;

π_2=.2440,μ_2=-6.1216,σ_2^2=.9713;

π_3=.2747,μ_3=1.9936,σ_3^2=1.6668;

π_4=.2338,,μ_4=6.0583,σ_4^2=.8504.

In this case the EM method is converging to the correct solution. We conclude our analysis of this example with a plot of the trajectories of the system in parameter space, see Figure 8. In this case we use a coordinate system given by (μ_x, π). The time evolution of the system is displayed by the four curves in the figure.

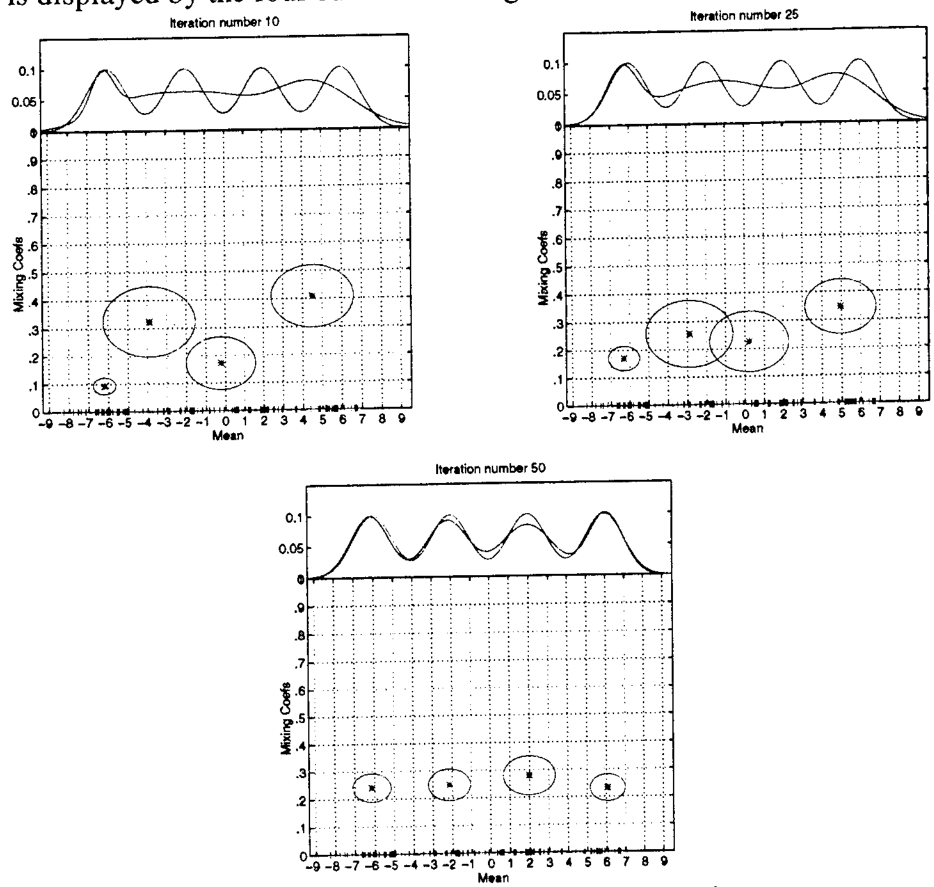


Figure 7: The system after (a) 10, (b) 25, and (c) 50 iterations.

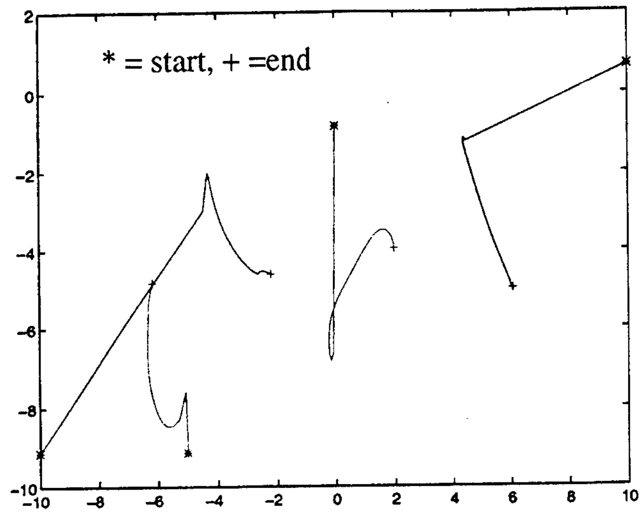


Figure 8: Phase space trajectories for the 4 term FMDE case.

We next turn our attention to a univariate case for the adaptive mixtures estimator. In this case our sample is 100 points drawn from $f_0(x)=.5N(-2,1)+.5N(2,1)$. Figure 9 displays the state of the system after the first data point. As promised, the model consists of a single term centered at this data point. Figures 9 b and c show the state of the system after the second and third data point. We notice that the second point fell within the support region of the first term and hence the model was updated using the recursive update equations and no term creation took place. A new term is created after the third point. Figures 10 a, b, and c show the state of the system after 25, 50, and 100 data points respectively. We notice the good fit between the adaptive mixtures model and the underlying probability distribution at this time.

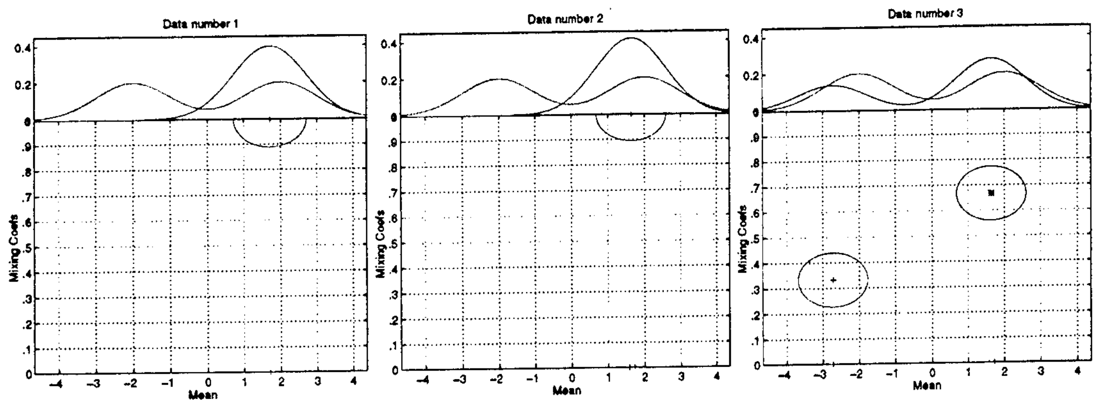


Figure 9: State of the adaptive mixtures procedure after (a) 1, (b) 2, and (c) 3 points.

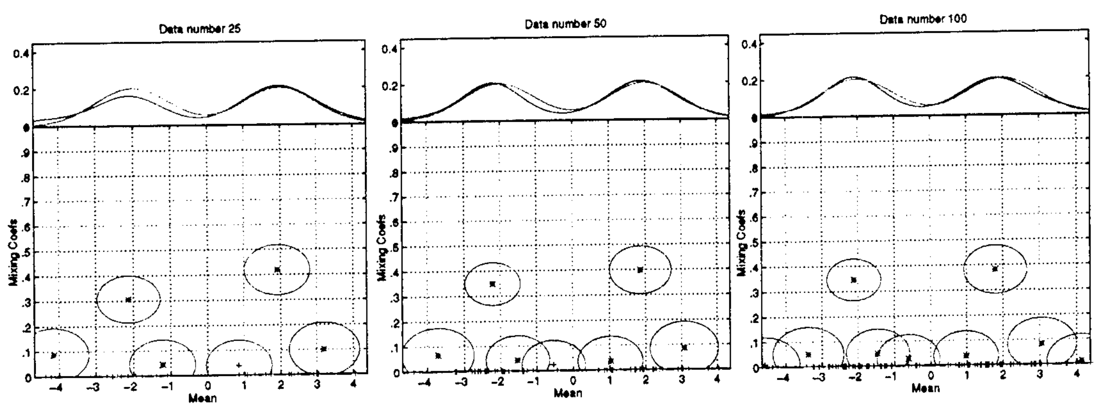


Figure 10: Adaptive mixtures solutions after(a) 25, (b) 50, and (c) 100 points.

We conclude our univariate examples by applying the adaptive mixtures procedure to a 100 data point set drawn from the four mode four term distribution of our first example. We present this to illustrate the different character of the solutions computed using the two procedures. Figure 11 illustrates the AMDE solution after the last data point. At this time, there is still a fairly good fit between the overdetermined mixture model and the true distribution. The overdetermined nature of the solution is a small price to pay when one considers that the model was produced without an initial estimate of the number of terms in the model or their position. In fact compared with the equivalent kernel estimator which contains 100 terms the AMDE is quite frugal.

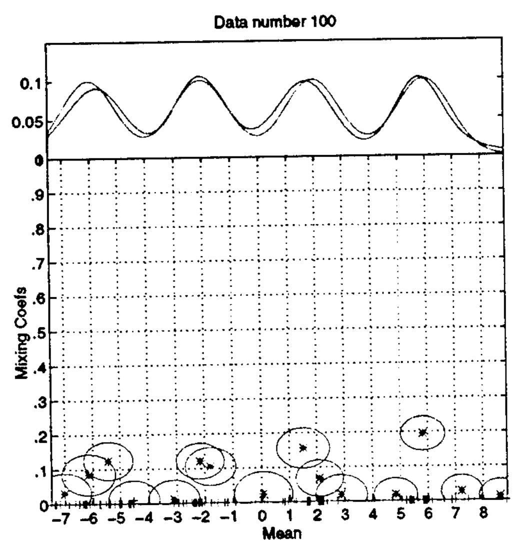


Figure 11: State of the AMDE after presentation of 100 points from a 4 mode 4 term distribution.

Bivariate Results

We next turn our attention to two bivariate examples. In the first one we consider

100 points drawn from a two component mixture given by

$$f_0(x, y) = 0.3N\left((-3, -3), \begin{bmatrix} 1. & 0.8 \\ 0.8 & 1. \end{bmatrix}\right) + 0.7N\left((3, 3), \begin{bmatrix} 1. & 0. \\ 0. & 1. \end{bmatrix}\right). \tag{13}$$

We first consider a two component finite mixtures solution based on this data set. The initial model is given by

$$f_0(x, y) = 0.5N\left((-5, -5), \begin{bmatrix} 1. & 0. \\ 0. & 1. \end{bmatrix}\right) + 0.5N\left((5, 5), \begin{bmatrix} 1. & 0. \\ 0. & 1. \end{bmatrix}\right) . \tag{14}$$

In Figures 12 a and b we present the initial configuration of the model and the model after 7 iterations through the data. We notice the close match between the final configuration of the mixture model and the true distribution..

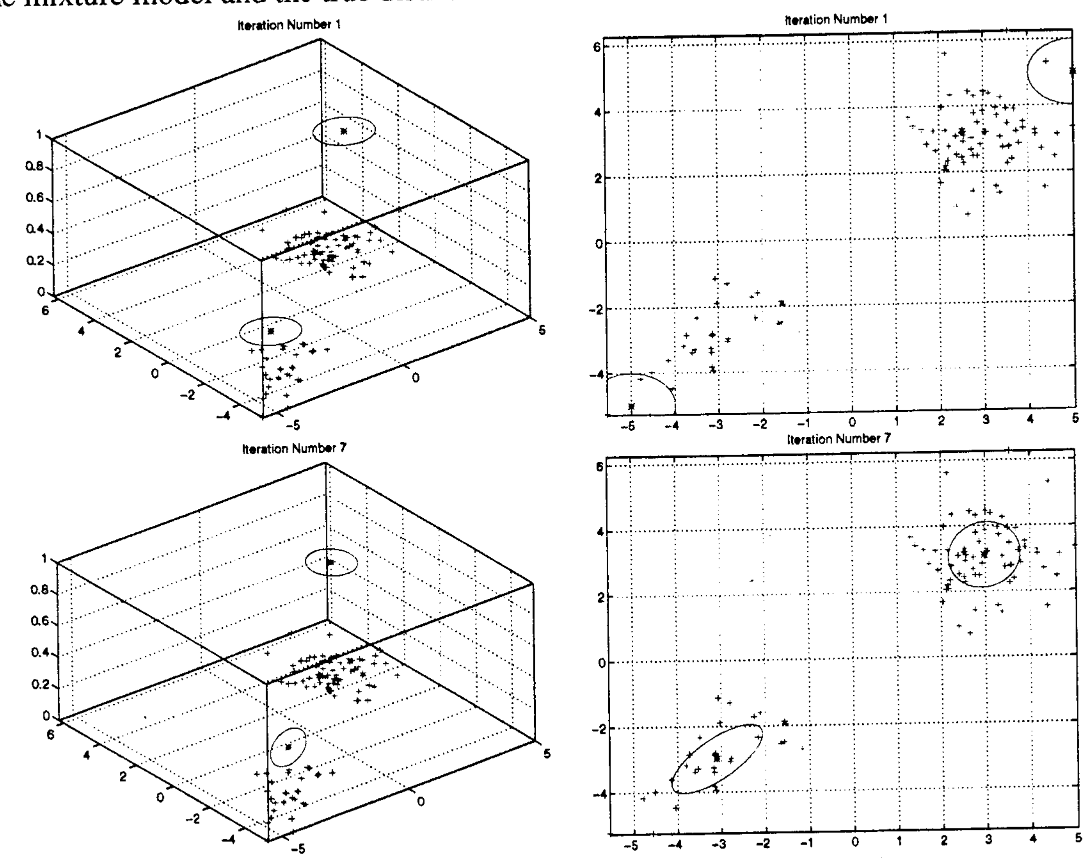


Figure 12: (a) Initial and (b) final configuration of two term finite mixture solution. View of [-37.5,45] and [0,90] have been presented in each case.

Once again it is interesting to compare the nature of the finite mixtures solution to that obtained using the adaptive mixtures procedure. In Figure 13 we portray the final configuration of the adaptive mixtures solution based on 100 points drawn from the above distribution. This solution which consists of 8 terms was obtained using a create threshold $\tau_c = 1.53^2 = 2.34$. This value was chosen to match the value of 1 used in the univariate simulations. We draw particular attention to the manner in which the estimator has modeled the leftmost region that contains the correlation. We see the terms placed end to end along the thin ridge. It is interesting to compare this with the long narrow term obtained by the finite mixtures estimator. Once again we see the utility of the visualization process

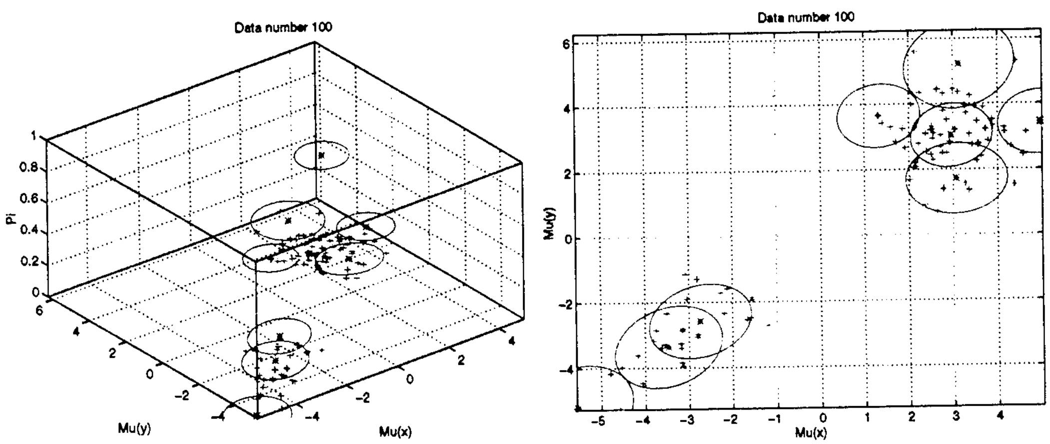


Figure 13: Final configuration of the eight term adaptive mixtures solution.

In Figures 14 a and b we present two views of the probability density function that results from this mixture. We see from view (a) that the relative heights of the two peaks seem

appropriate with respect the underlying mixing proportions. In view (b) we clearly see the correlated and uncorrelated peaks in the probability density function.

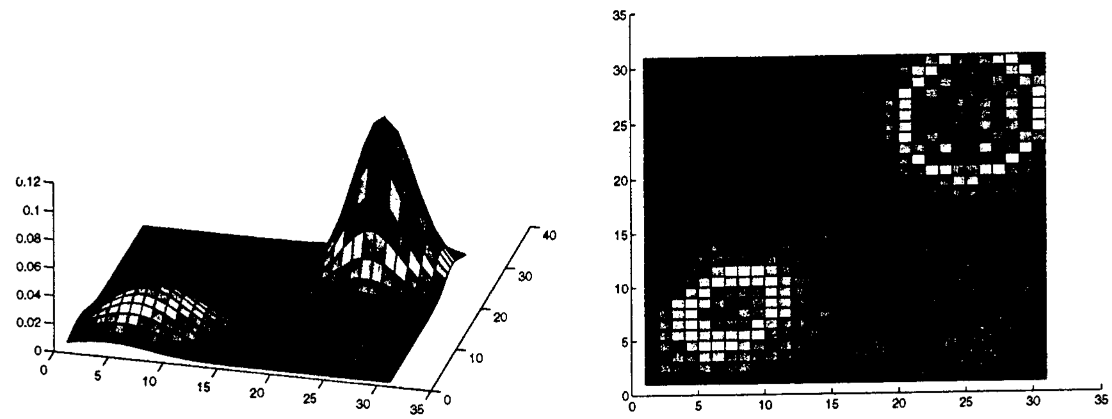


Figure 14: (a) and (b) Two views of the pdf corresponding to the solution of Figure 13.

Trivariate Results

The final case involves the adaptive mixtures analysis of a bimodal trivariate data set. The data set consists of 100 points drawn from

$$f_0(x, y, z) = 0.5N((-3,-3,-3), \Sigma_1) + 0.5N((3,3,3), \Sigma_2) \tag{15}$$

where Σ_1 and Σ_2 are given by $\Sigma_1 = \begin{bmatrix} 1.0 & 0.8 & 0 \\ 0.8 & 1.0 & 0 \\ 0 & 0 & 1.0 \end{bmatrix}$ and $\Sigma_2 = \begin{bmatrix} 1.0 & -0.8 & 0 \\ -0.8 & 1.0 & 0 \\ 0 & 0 & 1.0 \end{bmatrix}$. In Figures 15 a and b we present the solution after 25, and 100 points respectively. A create threshold of tc=1.88^2=3.54 was used. This value reflects the appropriate normalization for dimensionality. As indicated previously each term in the model is represented as a ellipsoid where the ellipsoid is determined by the covariance structure of the term. We notice the correlation structure of the data clearly indicated by the AMDE model.

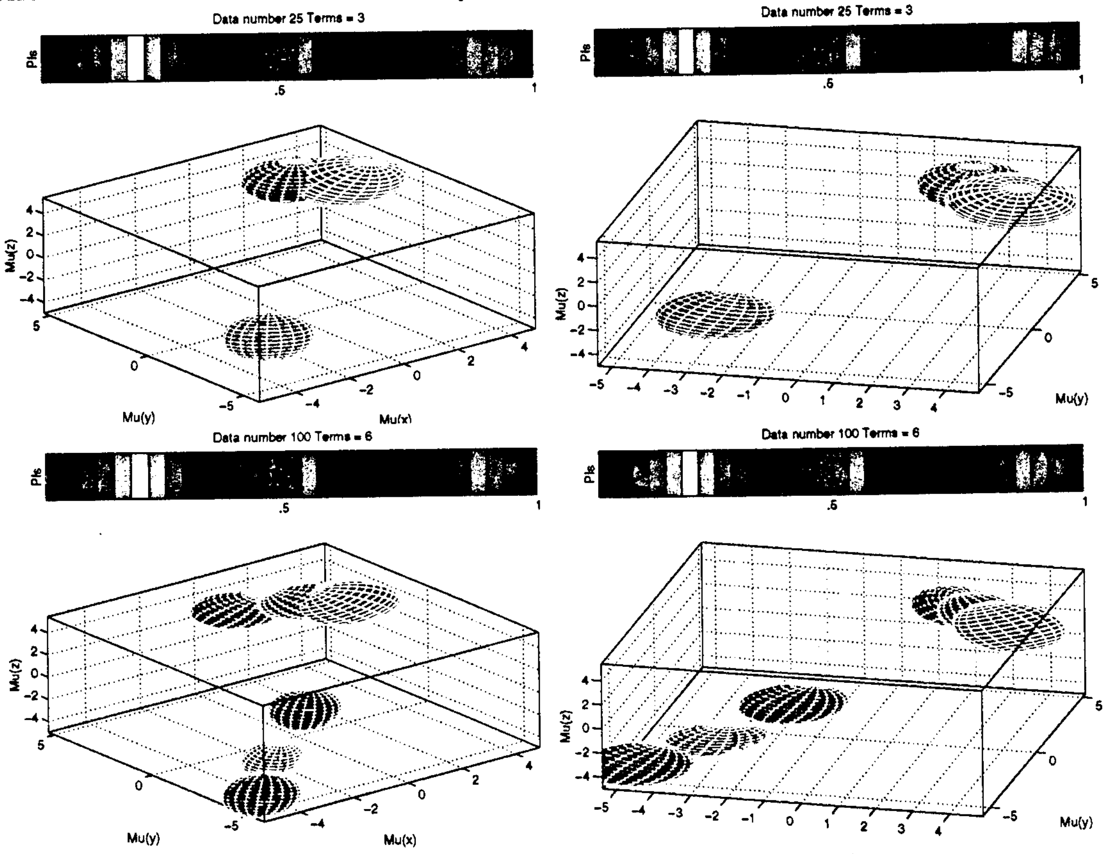


Figure 15: Trivariate AMDE solution after (a) 25, and (b) 100 points. Views of [-37.5,45] and [15,45].

On-line Access of Movies

Movies for each of these cases that we have discussed can be accessed via our on-line MOSAIC server at irisd.nswc.navy.mil (128.38.40.50). Some background discussion and mpeg movies are provided for each case. The reader is encouraged to view these movies in order to obtain a full appreciation of the process. In addition, the movies in MPEG and MATLAB format are available via anonymous FTP from irisd.

4: Conclusion

The EM algorithm can be used to perform maximum likelihood based estimation of unknown probability distributions. This estimation can take the form of the parametric finite mixtures procedure or the semi-parametric adaptive mixtures procedure. In either case, the time evolution of these systems can be very difficult to follow.

We have developed a new visualization technique to aid in the study of the time evolution of these parametric and nonparametric estimators in time. This technique makes use of graphical abstractions of the mixture model structure. We have found this procedure useful in gaining insights into the inner workings of both the finite mixtures and adaptive mixtures procedure.

We have also provided access to the movies produced using these visualization techniques. We plan to use these techniques to aid in our future research and pedagogical efforts. Some of our future research efforts will focus on the development of new adaptive bandwidth estimators that use alternate create criteria and estimation procedures.

5: Acknowledgments

The authors would like to thank the Office of Naval Research and the U. S. Marine Corps for the support of this research. In addition the authors would like to thank Dr. George Rogers of the Naval Surface Warfare Center and Dr. Carey Priebe of John Hopkins Universty for helpful discussions regarding dF space and the mixture visualization problem in general.

6: References

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm." *J. Royal Statist. Soc., Series B* 39 1-38.

Everitt, B. S. and Hand, D. J. (1981) *Finite Mixture Distributions.* Chapman and Hall. London, UK.

Priebe, C. E., Solka, J. L., and Rogers, G. W. (1993)"Discriminant Analysis in Aerial Images Using Fractal Based Features." in *Adaptive and Learning Systems II, F. A. Sadjadi, Ed., Proc. SPIE 1962*, pp. 196-208.

Priebe, C. E. (1994) "Adaptive Mixture Density Estimation." *Journal of the American Statistical Association*, vol. 89, No. 427, pp 796-806.

Priebe, C. E., Solka, J. L., Lorey, R. A., Rogers, G. W., Poston, W. L., Kallergi, M., Qian, W., Clarke, L. P., Clark, R. A. (1994) "The Application of Fractal analysis to Mammographic Tissue Classification." *Cancer Letters*, 77, pp. 183-189.

Priebe, C. E., Lorey, R. A., Marchette, D. J., and Solka, J. L. (1994) "Nonparametric Spatio-temporal Change Point Analysis for Early Detection in Mammography," presented at and appearing in the *Proceedings of the Second Int'l Workshop on Digital Mammography (SIWDM)*, York, UK.

Redner, R. A. and Walker, H. F.(1984), "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol 26, pp. 195-239.

Scott, D. W. (1985a) "Frequency Polygons." *J. Amer. Statist. Assoc.* 80, pp. 348-354.

Scott, D. W. (1985b) "Average Shifted Histograms: Effective Nonparametric Density Estimation in several Dimensions." *Ann Statist.* 13, pp. 1024-1040.

Scott, D. W. (1992) *Multivariate Density Estimation.* John Wiley and Sons, New York, NY.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall. New York, NY.

Solka, J.L., Priebe, C.E., and Rogers, G.W. (1992) "An Initial Assessment of Discriminant Surface Complexity for Power Law Features." *Simulation* Volume 58, No. 5, pp 311-318.

Solka, J. L., Priebe, C. E., and Rogers, G. W. (1993) "A Probabilistic Approach to Fractal Based Texture Discrimination." in *Adaptive and Learning Systems II, F. A. Sadjadi, Ed., Proc. SPIE 1962*, pp. 209-218.

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1994

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE
A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators

5. FUNDING NUMBERS
N00014-92-J-1303
N00014-93-1-0527

6. AUTHOR(S)
Jeffrey L. Solka, Wendy L. Poston
Edward J. Wegman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

8. PERFORMING ORGANIZATION REPORT NUMBER
no. 114

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Navy
Office of the Chief of Naval Research
Mathematical Sciences Division
800 N. Quincy Street Code 1111SP
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Navy position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper focuses on recent work which analyzes the expectation maximization (EM) evolution of mixtures based estimators. The goal of this research is the development of effective visualization techniques to portray the mixture model parameters as they change in time. This is an inherently high dimential process. Techniques are presented which portray the time evolution of univariate, bivariate, and trivariate finite and adaptive mixtures estimators. Adaptive mixtures is a recently developed variable bandwidth kernel estimator where each of the kernels is not constrained to reside at a sample location. The future role of these techniques in developing new versions of the adaptive mixtures procedure are also discussed.

14. SUBJECT TERMS
density estimation, expectation maximization, graphics

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Sturges, H. A. (1926) "The Choice of a Class Interval." *J. Amer. Statist. Assoc.* 21, pp. 65-66.

D. M. Titterington (1984) "Recursive parameter estimation using incomplete data." *J. Royal Stat. Soc., Ser. B, Vol. 46*, pp 257-267.

Wegman, E. J. (1970) "Maximum Likelihood Estimation of a Probability Density Function." *Sankhya. Ser. A, vl. 37,* pp. 211-224..

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1994

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE
A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators

5. FUNDING NUMBERS
DAAL03-91-G-0039
DAAH04-94-G-0267

6. AUTHOR(S)
Jeffrey L. Solka, Wendy L. Poston, Edward J. Wegman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

8. PERFORMING ORGANIZATION REPORT NUMBER
No. 114

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This paper focuses on recent work which analyzes the expectation maximization (EM) evolution of mixtures based estimators. The goal of this research is the development of effective visualization techniques to portray the mixture model parameters as they change in time. This is an inherently high dimential process. Techniques are presented which portray the time evolution of univariate, bivariate, and trivariate finite and adaptive mixtures estimators. Adaptive mixtures is a recently developed variable bandwidth kernel estimator where each of the kernels is not constrained to reside at a sample location. The future role of these techniques in developing new versions of the adaptive mixtures procedure are also discussed.

14. SUBJECT TERMS
density estimation, expectation maximization, graphics

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL